3D modeling and registration under wide baseline conditions

3D MODELING AND REGISTRATION UNDER WIDE BASELINE CONDITIONS

L. Van Gool[1,2], T. Tuytelaars[1], V. Ferrari[2], C. Strecha[1], J. Vanden Wyngaerd[1], and M. Vergauwen[1]

[1] ESAT/PSI/Visics, KULeuven, Belgium
[2] D-ITET/BIWI, ETH Zurich, Switzerland

KEY WORDS: wide baseline, 3D reconstruction, 3D registration, invariant neighbourhoods

ABSTRACT

During the 90s important progess has been made in the area of structure-from-motion. From a series of closely spaced images a 3D model of the observed scene can now be reconstructed, without knowledge about the subsequent camera positions or settings. From nothing but a video, the camera trajectory and scene shape are extracted. Progress has also been important in the area of structured light techniques. Rather than having to use slow and/or bulky laser scanners, compact one-shot systems have been developed. Upon projection of a pattern onto the scene, its 3D shape and texture can be extracted from a single image. This paper presents recent extensions on both strands, that have a common theme: how to cope with large baseline conditions. In the case of shape-from-video we discuss ways to find correspondences and, hence, extract 3D shapes even when the images are taken far apart. In the case of structured light, the problem solved is how to combine partial 3D patches into complete models, without a good initialisation of their relative poses.

1 INTRODUCTION

During the last few years, low-cost and user-friendly solutions for 3D modeling have become available. Shape-from-video (Armstrong 1994, Heyden 1997, Pollefeys 1998, Hartley 2000) extracts 3D shapes and their textures from video sequences as the only input. One-shot structured light techniques (Vuylsteke 1990, Proesmans 1996, Chia 1996, Eyetronics www) get such information from a single image, but need the projection of a special pattern. These techniques have the advantage that they are cheaper than traditional solutions like dedicated multi-camera rigs or laser scanners, as they only require off-the-shelf hardware. Moreover, they offer more flexibility in terms of portability and the range of object sizes they can handle.

This paper presents ongoing work on two different, but strongly related extensions of such systems.

Wide-baseline image matching: Shape-from-video requires large overlap between subsequent frames. Often, one would like to reconstruct from a small number of stills, taken from very different viewpoints. Based on local, viewpoint invariant features, wide-baseline matching is made possible, and hence the viewpoints can be farther apart.

Crude registration of 3D patches: Automatic registration algorithms for 3D patches such as ICP require good initial, relative positions and orientations of the patches to work. Completely automatic solutions to the 3D puzzle of putting together a set of unstructured 3D patches requires that a first, crude registration also takes place automatically.

2 WIDE-BASELINE IMAGE MATCHING

2.1 Task description

The 90s have witnessed the appearance of self-calibration techniques in structure-from-motion. A series of images is the only input such systems need to determine the camera motion and the evolution of the camera settings, as well as the 3D shape (up to an unknown scale) of the scene. By now, several approaches for such self-calibration have been developed and several systems have been proposed (Armstrong 1994, Heyden 1997, Hartley 2000, Pollefeys 1998). They start with the tracking of interest points through a sequence of views. The consistency of their image projections with a rigid 3D structure imposes constraints that allow to extract the cameras and the 3D shape of the cloud of interest points. The matching of these initial interest points will be referred to as *sparse correspondence* search. After the matching of the interest points, and the self-calibration, strong multi-view constraints between the images are available. These ease the search for many more correspondences. For one thing, a further search can be restricted to epipolar lines. In our approach (Pollefeys 1998), we go after pixelwise matches. This stage is referred to as *dense correspondence search*. These additional matches result in a detailed reconstruction of the 3D shape.

Although 3D reconstructions can in principle be made from a limited number of stills, these systems tend to only work effectively if the images have much overlap and are offered in the order of a continuous camera motion This is underlined by the name ‘shape-from-video’. For instance, we have tested our system (Pollefeys 1998) to make 3D records of archaeological, stratigraphic layers during excavations. A large part of the scene consists of sand and there is a general lack of points of interest. When walking around the dig, it proved necessary to take images less than 5^o apart. In such application, this is not always possible due to obstacles, and it disturbs the normal progress

Figure 1: *Two images of the same scene, but taken from very different viewing directions.*

of the excavations, as the image acquisition takes too much time, even when the images are taken in the form of a video sequence. It would be very advantageous, if the number of images can be limited to about 10 or so. These images would still cover the whole scene, but would be taken from substantially different viewpoints. Such 'wide baseline' images could also be taken with a digital photo camera rather than a video camera, leading to higher resolution imagery.

In summary, extending the shape-from-video technique to wide baseline conditions implies that both the sparse and the dense correspondence search have to be successful on images taken from very different viewpoints. The self-calibration procedure itself remains essentially identical. In our system, this is primarily based on the absolute quadric approach proposed by Triggs (Triggs 1997). Next, we describe the adapted versions of the correspondence steps.

2.2 Approach for sparse correspondence search

Consider the wide baseline image pair of fig. 1. The two images have been taken from very different viewing directions. Stereo and shape-from-video systems will most often not even get started in such cases, as correspondences are difficult to find.

As already mentioned, the shape-from-video approach splits the correspondence problem into two stages. The first stage determines correspondences for a relatively sparse set of features, usually corners. In the shape-from-video technique, the matching of corners is based on looking for corners within a region around the same position in the other image, and a selection on the basis of a normalised cross-relation of the surrounding intensity patterns. Both parts of this strategy will fail under the intended wide baseline conditions. The corresponding point may basically lie anywhere in the other image, and will not be found close to its original position. The use of simple cross-correlation will not suffice to cater for the change in corner patterns due to stronger changes in viewpoint and illumination. The next paragraphs describe an alternative strategy, that is better suited.

When looking for initial features to match, we should focus on local structures. Otherwise, occlusions and changing backgrounds will cause problems, certainly under wide baseline conditions. Here, we look at small regions, constructed around or near interest points. If these regions are to be matched, they ought to cover the same part of the scene in the different views. Hence, they have to take on different shapes in the different images. The most important aspect of the strategy proposed here is that the region extraction works on the basis of individual images, i.e. without any knowledge about the other images. This property is key to avoiding a slow and combinatoric search for matches. In the proposed scheme regions are constructed in one go based on a single image, instead of by selecting a region in one image and then trying to find a match by deforming and relocating a region in the other image until some matching score surpasses a threshold. Here, the corresponding region in the second image is extracted independently, before one even attempts to match regions. The crux of the matter is that every step in the region extraction is invariant under the image variations one wants to be robust against. This is discussed in more detail next.

On the one hand the viewpoint may strongly change. Hence, the extraction has to survive affine deformations of the regions, not just in-plane rotations and translations. In fact, affine transformations also not fully cover the observed changes. This model will only suffice for regions that are sufficiently small and planar. We assume that a reasonable number of such regions will be found, an expectation borne out in practice. On the other hand, strong changes in illumination conditions may occur between the views. The chance of this happening will actually grow with the angle over which the camera rotates. The relative contributions of light sources will change more than in the frame-to-frame changes in a video. We model the effects of changing illumination by scaling the three colour bands (R, G, B) with different scale factors and by adding different offsets. Our local feature extraction should also be immune against such photometric changes.

If we want to construct regions that are in correspondence

Figure 2: *'invariant neighbourhoods' that were extracted for the images in fig. 1. Only regions are shown for which a corresponding partner in the other image has been found, but the regions in the two images have been extracted without knowledge about the other image.*

irrespective of these changes and that are extracted independently, every step in their construction ought to be invariant under both the geometric and photometric transformations just described. A detailed description of these construction methods is out of the scope of this paper, and the interested reader is referred to papers specialised on the subject (Tuytelaars 1999, Tuytelaars 2000). As mentioned before, these constructions allow the computer to extract the regions in the two views completely independently. After they have been constructed, they can be matched efficiently on the basis of features that are extracted from the colour patterns that they enclose. These features again are invariant under both the geometric and the photometric transformations considered. To be a bit more precise, a feature vector of moment invariants is used. Fig. 2 shows some of the regions that have been extracted for fig. 1. We refer to the regions as 'invariant neighbourhoods'. Recently, several additional construction methods have been proposed by other researchers (Baumberg 2000, Matas 2001).

Also under the wide baseline version of shape-from-video, maybe better referred to as 'shape-from-stills', one is interested in finding correspondences between more than two

Figure 3: *Top row: views 1 and 2 of a bookshelf scene, with the 47 invariant neighbourhoods that have been matched indicated. Bottom row: the 41 matched invariant neighbourhoods for views 1 and 3 of the same scene.*

images. The previously described wide-baseline stereo matching approach is well suited for producing many feature matches between pairs of views that may be quite different. In practice, it actually is far from certain that the corresponding feature in another view will also be constructed by the system. Hence, the probability of extracting all correspondences for a feature in all views of an image set quickly decreases with the amount of views. Moreover, there is a chance of matching wrong features. For instance, let us suppose we are given 3 views v_1, v_2 and v_3. Although the method may find matches between the view pair $\langle 1, 2 \rangle$ and also between the view pair $\langle 1, 3 \rangle$, these two sets of matches will often substantially differ and a small number of common features between all three views may result. Figure 3 shows 3 views and the matches found between the pairs $\langle 1, 2 \rangle$ and $\langle 1, 3 \rangle$. Fig. 4 shows the matches that these pairs have in common. Whereas more than 40 matches were found between the pairs of fig. 3, the number of matches between all three views has dropped sharply, to only 16. When we consider 4 or 5 views, the situation can deteriorate further, and only a few, if any, features may be put in correspondence among all the views (even though there may be sufficient overlap between all the views).

Our most recent developments are devoted to counteract this problem. The approach is founded on two main ideas. Firstly, it is possible to exploit the information supplied by a correct match in order to generate many other correct matches. Suppose there is a feature A_1 in view v_1 which is matched to its corresponding feature A_2 in view v_2, and a feature B_1 in v_1 which could not get matched to its corresponding feature in v_2 (eg: the corresponding invariant neighbourhood B_2 has not been extracted, or maybe it has been extracted but the matching failed). If B_1 and A_1 are spatially close and lie on the same physical surface, then they will probably be mapped to v_2 by similar affine transformations. Hence, we can project B_2 in v_2 via the affine

Figure 4: *The features that could be matched in each of the 3 views of fig. 3. This intersection of the pairwise matching sets is quite small: only 16 features remain.*

transformation mapping A_1 to A_2 and get a first approximation of the real B_2. This approximation can then be refined by maximising the similarity between B_1 and the deformed B_2. We call this process *region propagation*. If B_1 is not close to A_1, or not on the same physical surface, a good similarity is unlikely to arise between the generated region and B_1, so this case can be detected and the propagated region rejected. The propagation approach strongly increases the probability that a feature will be matched between a pair of views, as it suffices that at least one feature in its neighborhood is correctly matched. As a result, also the probability of finding matches among all images of a set increases.

The second idea to obtain good quality multiview feature correspondences is to exploit redundant sets of matches between view pairs, or put differently, the transitivity property of valid matches. In our 3 view example, instead of only matching between the view pairs $\langle 1,3\rangle$ and $\langle 1,2\rangle$, we can also match 2 to 3. This introduces precious, additional information. For example, if a feature gets matched in $\langle 1,3\rangle$ but not in $\langle 1,2\rangle$, we can look if it is matched in $\langle 2,3\rangle$. If it is, at least one of these conclusions is wrong. Following a majority vote, we can conclude that the lack of a match in $\langle 1,2\rangle$ was a failure and obtain a correct feature correspondence along the three views.

In summary, starting from pairwise matches, many more can be generated. Of course, the validity of propagated and implied matches is an issue, and one has to be careful not to introduce erroneous information. More elaborated schemes to achieve this are the subject of a forthcoming paper, which currently is under preparation. The strategies proposed here are akin to recent work by Schaffalitzky and Zisserman (Schaffalitzky 2002). In contrast to their work, there is less emphasis on computational efficiency. In particular, adding transitivity reasoning to the propagation of matches renders our approach slower, but it also adds to the performance. The combined effect of propagation and transitivity reasoning for our example is illustrated in fig. 5. The number of matches along the three views has more than tripled.

2.3 Approach for dense correspondence search

The matching of invariant neighbourhoods is only the first step in the search for correspondences. Good 3D models require the selection of dense, pixelwise correspondences. In the shape-from-video pipeline, the initial, sparse corner matches provide epipolar constraints, that simplify the subsequent dense correspondence search. Within this wide baseline setting, it are the invariant neighbourhoods which provide the epipolar constraints. But also with these constraints in place, dense correspondence search under wide baseline conditions requires adaptations. Although our current dense correspondence algorithm (Van Meerbergen 2002), which is based on a kind of dynamic path search along epipolar lines, performs quite well under changes that are a bit larger than the ones between subsequent video

Figure 5: *The features that could be matched in each of the 3 views of fig. 4 after propagation and transitivity reasoning. The number of matches has been increased to 58.*

frames, it nevertheless has difficulties coping with more extreme cases.

Under wide baseline conditions, disparities tend to get larger, a smaller part of the scene is visible to both cameras, and intensities of corresponding pixels vary more. In order to better cope with such challenges, we propose a scheme that is based on the coupled evolution of Partial Differential Equations. This approach is described in more detail in a paper by Strecha *et al.* (Strecha 2002). The point of departure of this method is a PDE-based solution to optical flow, proposed earlier by Proesmans *et al.* (Proesmans 1994). In a recent benchmark comparison between different optical flow techniques, this method performed particularly well (McCane 2001). An important difference with classical optical flow is that the search for correspondences is 'bi-local', in that spatio-temporal derivatives are taken at two different points in the two images. Disparities or motions are subdivided into a current estimate and a residue, which is reduced as the iterative process works its way towards the solution. This decomposition makes it possible to focus on the smaller residue, which is in better agreement with the linearisation that is behind optical flow. The non-linear diffusion scheme in the Proesmans *et al* approach imposes smoothness of nearby disparities at most places – an action which can be regarded as the dense counterpart of propagation – but simultaneously allows for the introduction of discontinuities in the disparity map.

The method of Strecha *et al.* (Strecha 2002) generalises this approach to multiple views. The extraction of the different disparities is coupled through the fact that all corresponding image positions ought to be compatible with the same 3D positions. The effect of this coupling can be considered the dense counterpart of the sparse transitivity reasoning. Moreover, the traditional optical flow constraint that corresponding pixels are assumed to have the same intensities, is relaxed. The system expects the same intensities *up to scaling*, where the scaling factor should vary smoothly between neighbouring pixels at most places.

2.4 Experiments

Fig. 6 shows three images of the left corner of the town hall of Leuven. These images are too far apart for our shape-from-video process to get started with the corner matching. A sufficient number of invariant neighbourhoods can be matched, however, and the PDE-based dense correspondence search succeeds in finding matches for most other pixels. Three views of the resulting 3D model are shown in fig. 7. The result looks quite convincing, even for such a convoluted surface, where parts easily get occluded in several views. This problem of holes in the model precluded us from taking the images even farther apart.

Fig. 8 shows three images of an excavation layer, acquired at the Sagalassos site in Turkey. This is one of the largest scale excavations currently ongoing in the Mediterranean, under the leadership of prof. Marc Waelkens. These im-

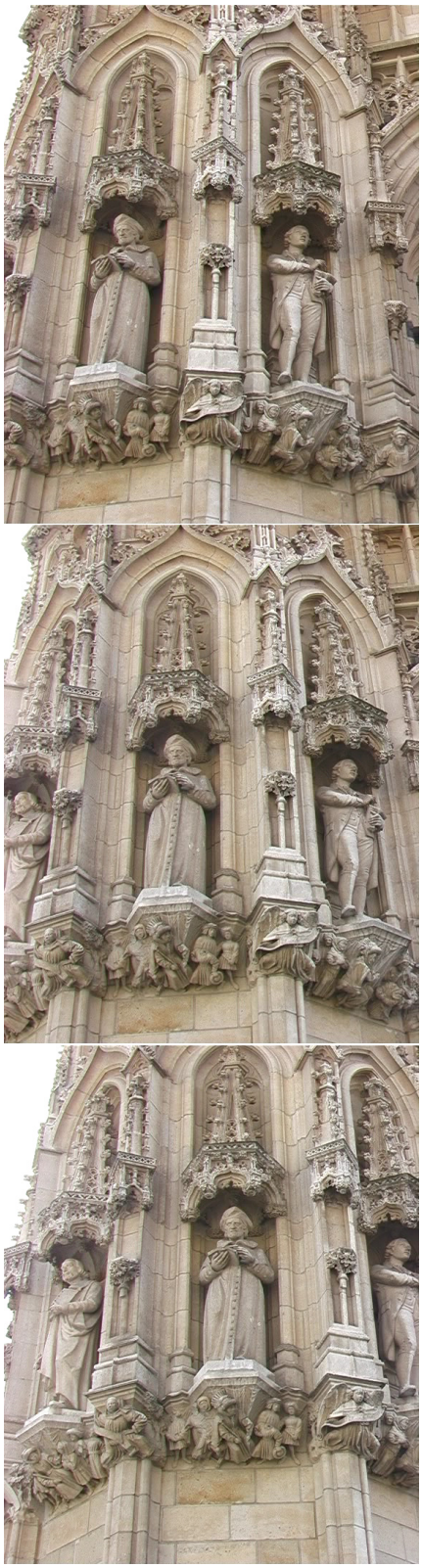


Figure 6: *Three input images of the ornamental facade of the town hall in Leuven. The images are too far apart for our shape-from-video process to match features between the views.*

Figure 7: *Three reconstructions extracted from the relatively wide baseline images of fig. 6.*

Figure 8: *Three input images of an excavation layer at an archaeological site. The images are too far apart for our shape-from-video process to match features between the views.*

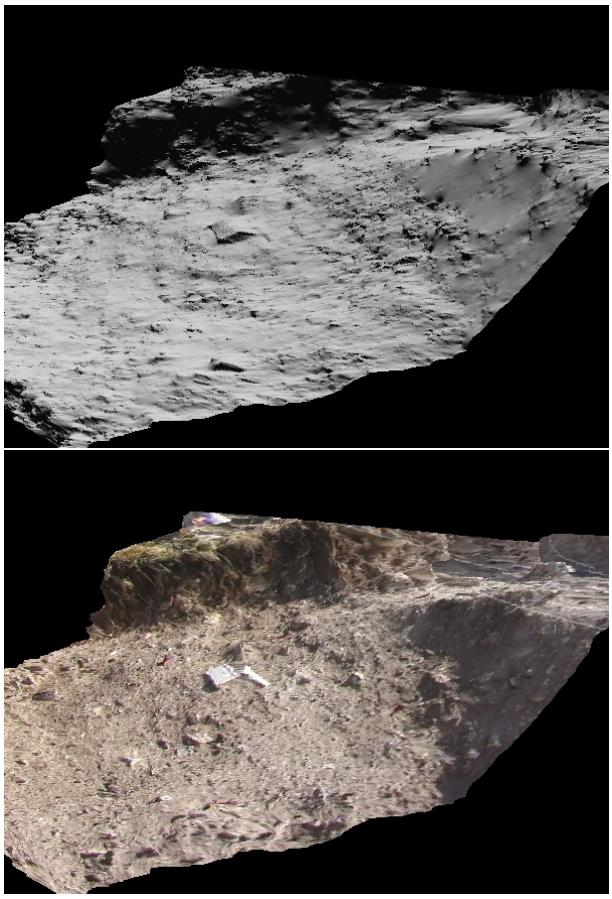


Figure 9: *The reconstruction extracted from the relatively wide baseline images of fig. 8, with and without texture.*

ages have less structure than the ones of the town hall and are too far apart for our shape-from-video process to get its corner matching started successfully. Again, invariant neighbourhoods haven been matched and the PDE-based dense correspondence search succeeded in finding matches for most other pixels. A side view of the resulting 3D model is shown in fig. 9, with and without the texture.

3 AUTOMATIC, CRUDE REGISTRATION OF 3D PATCHES

3.1 Task description

If partial 3D reconstructions have already been produced from different photo sets, model completion may better be done in 3D. Similarly, there are a new generation of structured light techniques that generate partial, 3D patches from each picture that is taken. If there is sufficient overlap between the 3D patches, they can be fitted together to build a single, complete shape model. This fitting together of patches is usually referred to as 'registration'.

The state-of-the-art in 3D registration is similar to that in 2D. Several, excellent methods have been proposed to precisely fit together partial, 3D reconstructions from initial

positions that are almost in correct relative positions (Besl 1992, Chen 1991, Viola 1995). Distances between corresponding points on different patches are small in that case. Such automatic fine registration is very important, as it is usually easier to manually position the 3D patches more or less right, than it is to perform the fine docking by hand. Of course, it would be nicer if also the initial, crude positioning could be done by the computer, as this would render the whole registration automatic. Also, if the 3D patches are presented to the system as an unstructured set, it will in many cases be difficult to find out which patches would fit. The problem becomes a 3D puzzle. Performing also crude registration automatically is the very goal of the work described next.

Much like with a normal, 2D puzzle the pieces can be put together on the basis of two complementary types of information. On the one hand there is their shapes, which should match. Here, it is not a matter of outlines that should tally, but the 3D shapes ought to be the same for the part where the patches overlap. On the other hand, the surface may contain texture. If this is captured by the structured light system, then it may yield very effective clues as well. Even more so, in cases where the shape does not allow to build an unambiguous reconstruction for reasons of symmetry, the texture may break this symmetry.

3.2 Approach for geometry-based registration

Assume one has to find matches between overlapping, 3D patches. These patches overlap only partially. A naive way to approach the problem would be to take any pair, and to search for a Euclidean motion in 3D that generates a good fit. This process would be prohibitively slow.

Again, invariants have proven instrumental in the development of methods that achieve such crude registration from arbitrary, initial 3D patch positions. They use special points or curves on the surface, which are characterised with invariants (Feldmar 1994, Johnson 1997). A feature type that we have found to be particularly useful are bitangent curves. They are interesting, because they are invariant under Euclidean, affine, and even projective transformations. Moreover, the curve pairs can be given simple, invariant descriptions, especially in the case of Euclidean and affine transformations. These descriptions require only first derivatives (Vanden Wyngaerd 1999). Bitangent curves are formed as follows. Suppose a plane touches the surface at two points (i.e. it is a 'bitangent plane'). Now one rolls this plane over the surface so that it keeps in touch at two points. This yields pairs of bitangent curves, as illustrated in figure 10.

For the computation of bitangent curves we construct a dual surface. Rothwell (Rothwell 1994) already used dual representations of planar curves to find pairs of bitangent points. In that case, the dual is a curve and a bitangent point pair corresponds to a self-intersection of the dual. Here we use a direct extension of this idea for surfaces.

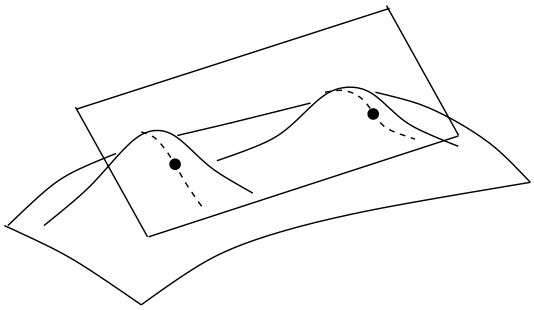
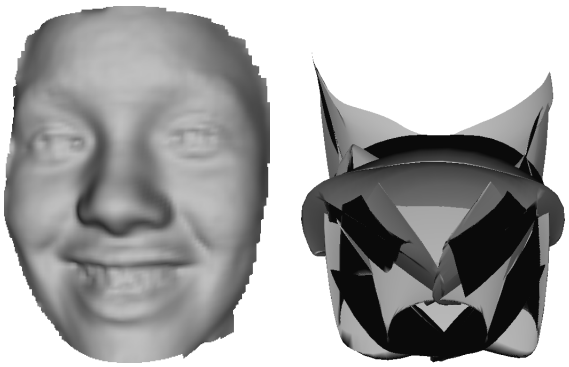


Figure 10: *Bitangent planes can roll over the surface, thereby describing pairs of bitangent curves.*

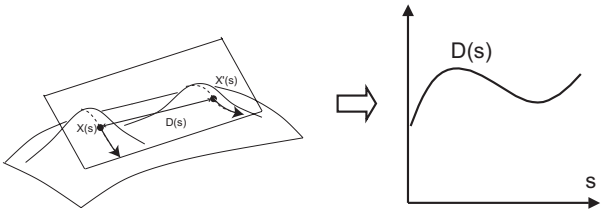
For every point $\mathbf{X}$ of the surface the tangent plane is calculated. This tangent plane can be represented by three parameters. These three parameters are used to create a three-dimensional dual point. Replacing all surface points by their dual results in a dual surface. Since bitangent points have the same tangent plane, they have the same dual point and the bitangent curve pairs correspond to curves of self-intersection of the dual surface. Figure 11 shows an example of such a dual surface.



(a) (b)

Figure 11: *(a) An example surface. (b) View of its dual surface. The dual surface is constructed by replacing all points by their duals. As described in the text, the dual point of a surface point* $\mathbf{X}$ *represents its tangent plane.*

In our approach, crude registration is carried out through the matching of bitangent curve pairs on the different patches. In order to support efficient curve matching and to find point correspondences between different patches, we use an invariant description of the bitangent curve pairs. In our patch registration problem, invariance under 3D rotation and translation suffices. The bitangent curves are characterized by invariant signatures, which express an invariant as a function of an invariant parameter. A problem with signatures of single space curves is that they may require higher derivatives, such as the 2nd and 3rd derivatives for curvature and torsion in the Euclidean case (Mokhtarian 1997). Semi-differential invariants (Van Gool 1992) use lower order derivatives in more than one point. Pajdla and Van Gool (Pajdla 1995) used them for Euclidean registration of space curves. In general, semi-differential invariants use fixed reference points in combination with a vary-



(a) (b)

Figure 12: *A bitangent point pair* $(\mathbf{X}, \mathbf{X}')$ *slides along the curve pair. The same parameterization can be used for both curves which simplifies calculation of invariants. We compute the distance between the bitangent point pair* $(\mathbf{X}, \mathbf{X}')$ *as it slides along the curves. (b) An invariant signature of a bitangent curve pair expresses the distance between the points as function of the arclength of the longest curve of the pair.*

ing point. This introduces the problem that an expression is only invariant if the same reference point is used. In the case of bitangent curve pairs, the reference point can be the corresponding bitangent point on the other curve, thereby further simplifying the construction of stable invariant signatures. Hence, two points are combined that slide together along the bitangent curves.

In the Euclidean case the distance between bitangent points is invariant. The Euclidean arclength serves as an invariant parametrization. A single parametrization is used for both curves, for which we use the arclength of the longest curve of the pair. By computing the distance as the bitangent point pair slides along the curves, we get an invariant signature as illustrated in figure 12. These signatures are well suited for the matching of curves found on different patches. Rather than trying to directly match different surface patches, the goal is to match their most salient bitangent curves. In order to render the process more efficient and more robust, the search for matching signatures starts with the 15 *longest* bitangent curves. For efficiency, this length is measured as the number of sample points of the self-intersection curve in the dual space. Only these curves will be converted into bitangent curve pairs. This procedure does not exactly select the bitangent curves that have the longest arclength, but it gives a fair approximation.

The matching can be done efficiently by matching their invariant signatures. As a criterion, we use the $L2$-norm. As it may very well happen that only parts of bitangent curves are found on each of the patches, the signatures are divided into segments of equal length and these segments are matched. For efficient comparison, the signatures on different surfaces are resampled with the same constant arclength between sample points. Finding the best matching segments between two signatures is done by having a segment of the first signature slide along the second. No guarantee exists that corresponding bitangent curve pairs on different surfaces are parameterized in the same direction. This means that the starting point and ending point of the signatures can be inverted. We take this possibility into account in the matching process by checking whether the signature increases or decreases over the segment.

A signature segment as defined in the previous paragraph corresponds to two 3D curve segments, one on each curve of the bitangent curve pair. Consequently, its endpoints define four points on the surface patch. If a pair of signature segments is matched successfully, this suggests a match between 4 points on the two surface patches. These typically yield enough information to obtain a crude estimate for the transformation between the patches. Every matching signature segment provides us with a candidate transformation. Signature matches are ranked according to their $L2$-norm. However, only looking at the $L2$-norm does not suffice to select the best transformation candidate because signatures can match exactly without corresponding to the correct transformation. A typical example is a left-right symmetric face. Bitangent curve pairs will be symmetric, and signature segments from the left side can be matched exactly with the ones on the right side. The transformation implied by left-right mismatched signatures will result in noses pointing in the opposite direction. In order to eliminate these ambiguities, a verification step is done on the best matching signature segments. After a good signature match is found, it is checked by applying the corresponding transformation and by verifying how well the surfaces 'fit'.

3.3 Approach for texture-based registration

Wide baseline matching between 3D patches can also be based on their surface texture, rather than their 3D shape. A direct extension of the previous ideas would be to extract intensity or colour edges in the texture maps, which correspond to space curves on the patch surfaces. Such curves can then be matched, as e.g. proposed by Pajdla and Van Gool (Pajdla 1995). Here we follow a different and more robust strategy. The invariant neighbourhoods are used again. This makes us less dependent on the presence and clean extraction of edges in the texture maps. This also renders the features more local and therefore better suited for cases with limited overlap between patches. The invariant neighbourhoods can cope with the deformations that may exist between different texture maps. The actual matching is simple then. Invariant neighbourhoods are extracted from the texture maps of the patches, are matched based on their feature vectors of moment invariants, and from each of the successfully matched neighbourhoods a few points are selected (e.g. the center point of the neighbourhood). Next, a 3D Euclidean motion is determined that minimises the sum of distances between the corresponding points of corresponding neighbourhoods. This transformation is computed with Horn's quaternion based method (Horn 1987).

3.4 Experiments

A first example shows the matching of 3D patches based on shape. The patches are shown in fig. 13 and belong

to the well-known Stanford bunny, typically used as a demo object by the computer graphics community (Stanford 3D Scanning Repository). The bitangent curves of these patches were extracted and then matched based on their invariant signatures. Fig. 14 shows the bitangent curves for the first view. The automatically registered bunny data is shown in fig. 15. Note that this completed model is the result of the automated 'crude registration'. A fine registration based on ICP or another technique can be used to refine it. Nevertheless, it already looks quite convincing. The automatic matching (incl. bitangent curve extraction) took about 9 min. on a Pentium III 1.1 GHz.

A second example illustrates 3D patch matching on the basis of the texture maps, i.e. on the basis of invariant neighbourhoods extracted from these. It goes without saying that for this technique to be useful the 3D acquisition device should also capture the surface texture. We have used Eyetronics' ShapeWare (Eyetronics www). We demonstrate our approach for a globe. This is an example where a shape-based approach is doomed to fail, due to the high degree of shape symmetry. The texture with a representation of the continents and oceans breaks this symmetry and makes it possible to automatically arrive at a complete compilation of the object model. Fig. 16 shows two of the 48 patches that were captured separately. As can be seen, the overlap is rather small. Yet, more than 200 corresponding invariant neighbourhoods could be found (without propagation and transitivity reasoning in this case). A detailed cutout of both patches with the matching neighbourhoods is shown in fig. 17. The globe could be reconstructed automatically based on the texture approach alone. A view of the result is shown in fig. 18. Just as in the case of shape-based registration, it is advisable to apply a texture-based fine registration after this rather crude stage. Johnson and Kang have proposed an approach that could serve this purpose (Johnson 1999). This second stage should then also take care of texture blending. As can be seen in fig. 18 the original patches in our reconstruction can still be distinguished by their differences in texture intensities.

4 CONCLUSIONS AND FUTURE WORK

Three-dimensional reconstruction often introduces 'wide baseline' problems. This can be the case at the point where one has to find correspondences between the 2D input views, or when one has to register partial 3D reconstructions. We have proposed solutions to both 2D and 3D wide baseline matching problems. Ongoing work is mainly focused on issues of efficiency. A stronger integration of 2D and 3D techniques remains to be explored.

Acknowledgements: The authors gratefully acknowledge support from K.U.Leuven GOA project 'VHS+' and European IST project '3D-MURALE'. Help by prof. Marc Waelkens and K. Cornelis of the Kath. Univ. of Leuven in gathering the archaeological imagery is gratefully acknowledged. We also thank J. Matas for providing images.

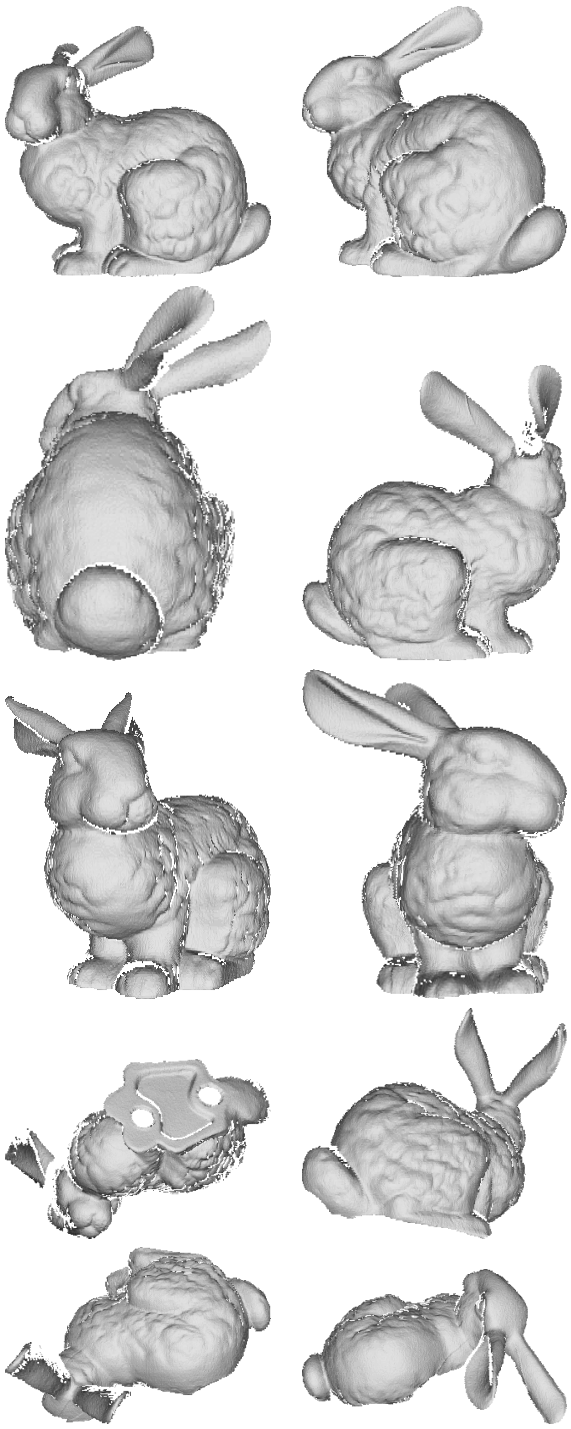


Figure 13: *Range data from the Stanford bunny.*

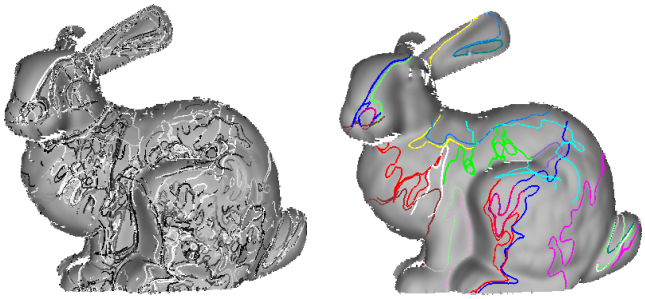


Figure 14: *Bitangent curve pairs on the first patch of the bunny. (a) 3072 curve pairs are detected. (b) Here we only the 15 longest curve pairs. These are the ones that are used for matching.*

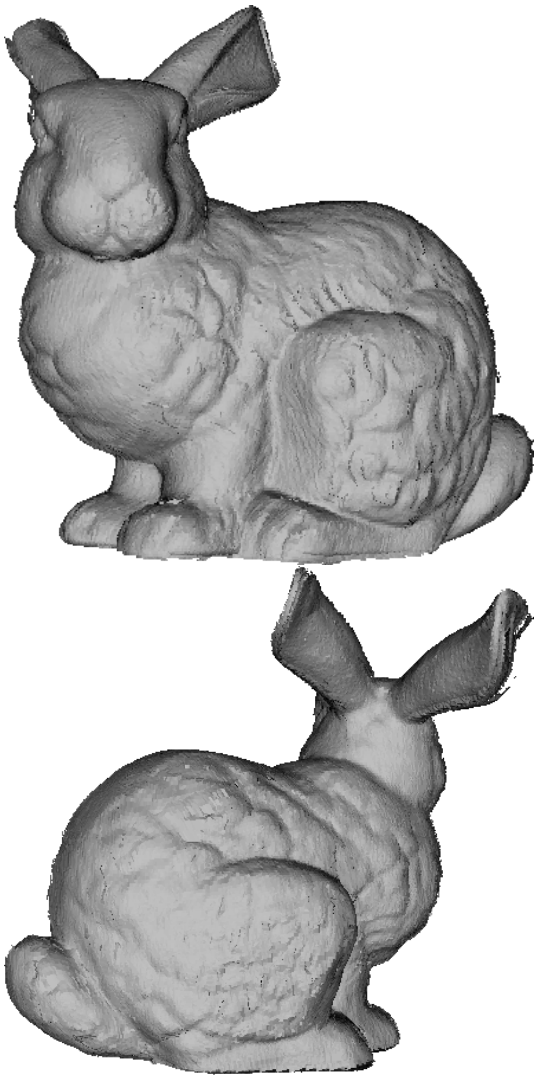


Figure 15: *Two views on the Stanford bunny after bitangent based crude registration of the range data in fig.13.*

Figure 16: *Two patches of a globe out of a total of 48, that were acquired in 3D separately.*

Figure 17: *Detailed cutouts of the two patches shown in fig. 16. The invariant neighbourhoods that could be matched in these cutouts are highlighted.*

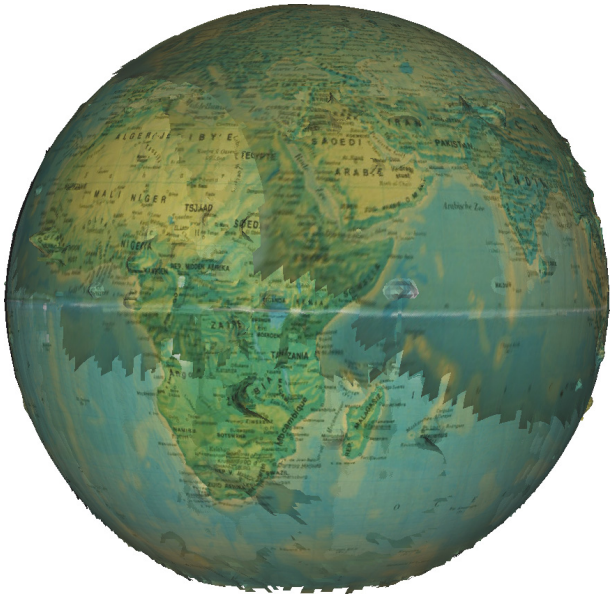


Figure 18: *View on the automatically completed globe model. Only texture information was used. In fact, shape would not suffice for this highly symmetric object.*

REFERENCES

M. Armstrong, A. Zisserman, and P. Beardsley, *Euclidean structure from uncalibrated images*, 5th BMVC, 1994

A. Baumberg, *Reliable Feature Matching Across Widely Separated Views*, Proc. CVPR, pp. 774-781, 2000

P. Besl, N. McKay, *A method of registration of 3-D shapes*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(2):239-256, 1992

Y. Chen and G. Medioni, *Object modeling by registration of multiple range images*, Proc. Int. Conf. on Robotics and Automation, pp. 2724-2729, 1991

T. Chia, Z. Chen, and C. Yueh, *Curved surface reconstruction using a simple structured light method*, Proceedings of the International Conference on Pattern Recognition, Vol. A, pp. 844-848, 1996

Eyetronics, http://www.eyetronics.com

J. Feldmar and N. Ayache, *Rigid, affine and locally affine registration of free-form surfaces*, TR INRIA Epidaure, No. 2220, 1994

R. Hartley and A. Zisserman, *Multiple View Geometry*, Cambridge University Press, 2000

A. Heyden and K. Astrom, *Euclidean reconstruction from image sequences with varying and unknown focal length and principal point*, Proc. CVPR, 1997

Berthold K. P. Horn, *Closed-form solution of absolute orientation using unit quaternions*, Journal of the Optical Society of America A, 4(4):629-642, April 1987.

A. Johnson and M. Hebert, *Recognizing objects by matching oriented points*, Proc. CVPR, pp. 684-689, 1997

A. Johnson and S. Kang, *Registration and integration of textured 3D data*, Image and Vision Computing, 17, pp. 135-147, 1999

J. Matas, O. Chum, M. Urban, and T. Pajdla, *Distinguished regions for wide-baseline stereo*, Research Report CTU–CMP–2001–33, Center for Machine Perception, Czech Techn. Un., Prague, November 2001.

B.McCane, K.Novins, D.Crannitch and B.Galvin, *On Benchmarking Optical Flow*, Computer Vision and Image Understanding, 84(1):126-143, 2001.

F. Mokhtarian, A Theory of Multiscale, *Torsion-Based Shape Representation for Space Curves*, Computer Vision and Image Understanding, 68(1):1-17, October 1997.

T. Pajdla and L. Van Gool, *Matching of 3-D curves using semi-differential invariants*, Proc. ICCV, pp. 390-395, 1995.

M. Pollefeys, R. Koch, and L. Van Gool, *Self calibration and metric reconstruction in spite of varying and unknown internal camera parameters*, Proc. ICCV, pp.90-96, 1998

M. Proesmans, L. Van Gool, E. Pauwels, and A. Oosterlinck, *Determination of optical flow and its discontinuities using non-linear diffusion*, Proc. ECCV, Stockholm, pp. 295-304, May 1994

M. Proesmans, L. Van Gool, and A. Oosterlinck, *One-shot active 3D shape acquisition*, Proceedings of the International Conference on Pattern Recognition, pp.336-340, 25-29 Augustus 1996, Vienna, Austria

C. Rothwell, *Recognition using projective invariance*, Ph.D. Thesis, Oxford University, 1994.

F. Schaffalitzky and A. Zisserman, *Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?"*, Proc. ECCV, pp. 414-431, Copenhagen, 2002

C. Srecha and L. Van Gool, *PDE-based multi-view depth estimation*, Proc. 1st Int. Symp. on 3D Data Processing, Visualization, and Transmission (3DPVT), pp. 416-425, Padova, june 19-21, 2002

The Stanford 3D Scanning Repository, http://www-graphics.stanford.edu/data/3Dscanrep/

W. Triggs, *Auto-calibration and the absolute quadric*, Proc. CVPR, pp. 609-614, 1997

T. Tuytelaars, L. Van Gool, L. D'haene, R. Koch, *Matching Affinely Invariant Regions for Visual Servoing*, International Conference on Robotics and Automation, Detroit, pp. 1601-1606, May 10-15, 1999

T. Tuytelaars and L. V-an Gool, *Wide baseline stereo matching based on local, affinely invariant regions*, Proc. British Machine Vision Conference, Vol. 2, pp. 412-425, Bristol, 11-14 sept, 2000

L. Van Gool, T. Moons, E. Pauwels, and A. Oosterlinck, Semi-differential invariants, in *Applications of invariance in vision*, eds. J. Mundy and A. Zisserman, pp. 157-192, MIT Press, Boston, 1992

G. Van Meerbergen, M. Vergauwen, M. Pollefeys, and L. Van Gool, *A hierarchical symmetric stereo algorithm using dynamic programming*, International Journal of Computer Vision, 47(1-3):275-285, 2002

J. Vanden Wyngaerd, L. Van Gool, R. Koch, and M. Proesmans, *Invariant-based registration of surface patches*, Proc. ICCV, pp. 301-306, Kerkyra, Greece, 1999

P. Viola and W. Wells, *Alignment by maximisation of mutual information*, Proc. ICCV, pp. 16-23, 1995

P. Vuylsteke and A. Oosterlinck, *Range Image Acquisition with a Single Binary-Encoded Light Pattern*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(2):148-164, 1990